ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-4485-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICASSP48485.2024.10446917

ROFI: ROBUST WIFI INTRUSION DETECTION VIA DISTRIBUTION MATCHING

*Xu Wang, Dongheng Zhang**, *Fengquan Zhan, Xuecheng Xie, Pengcheng Huang, Yang Hu, Yan Chen*

School of Cyber Science and Technology, University of Science and Technology of China, Hefei, China

ABSTRACT

Intrusion detection acts as a key to in-home security, where WiFi-based systems have gained wide attention due to the ubiquitous nature of WiFi signals. While existing methods achieve impressive performance in specific environments, they are susceptible to environmental changes, especially for complex scenarios where outdoor human activities can be mistaken as intrusions. In this paper, we propose RoFi, a robust WiFi intrusion detection system which can handle more complex scenarios. It achieves this by exploring the distribution of autocorrelation function (ACF) of Channel State Information (CSI) when intrusion occurs, where likelihood ratio testing is employed to discriminate intrusion and non-intrusion scenarios, eliminating the variance of different environments. Without complex calibration, RoFi achieves an accuracy of over 97.5% in practical deployment, outperforming existing methods.

Index Terms— Indoor Intrusion Detection, Wireless Sensing, Channel State Information (CSI)

1. INTRODUCTION

Indoor safety has always been a paramount concern for people [1] in diverse environments including domestic and public spaces. As an important tool to provide timely alerts when intruders appear indoors, intrusion detection systems can help people quickly understand the situation and avoid serious incidents, which enables applications including asset protection, in-home security, and child care.

However, existing intrusion detection methods based on vision or other sensors suffer from the inherent drawback of the modalities they rely on. Vision-based methods [2] are vulnerable under low-light and smoky environments. The privacy risk has further restricted the application of such systems. Sensor-based methods [3, 4] usually require purchasing specialized sensors and deploying numerous nodes, leading to high cost and labor-intensive deployments.

With the prevalent application of WiFi for in-home environments, enabling ubiquitous sensing with WiFi has gained wide attention. Applications such as indoor positioning [5, 6], target tracking [7], intrusion detection [8, 9], fall detection [10, 11], respiration monitoring [12, 13, 14] has been achieved using commodity WiFi devices.

Corresponding author: Dongheng Zhang. This work was supported by Anhui Key R&D Programmes under Grant 2022h11020026, National Key R&D Programmes under Grant 2022YFC2503405 and 2022YFC0869800.

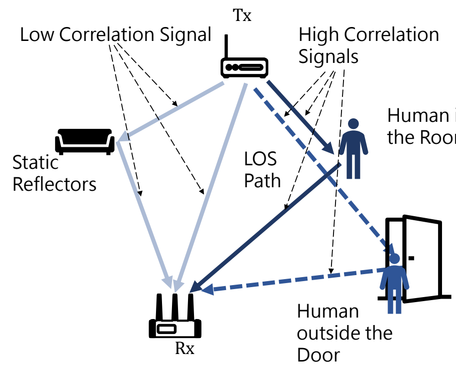


Fig. 1. The Propagation of WiFi Signals.

Despite many advantages provided by WiFi-based intrusion detection, existing methods still suffer from performance degradation in complex scenarios. For example, the measurement in WiBorder [8] works well when the Tx and Rx locations are fixed, but needs threshold adjustment when the device placement changes. While WiDetect [9], with a large coverage area, may not effectively distinguish between movements outside and inside the room.

To address the limitations of existing intrusion detection methods, we propose RoFi, a new approach for robust WiFi-based intrusion detection. Inspired by the idea of modeling WiFi signals as time series [15], we take a further step to characterize the autocorrelation function (ACF) of Channel State Information (CSI) power as Gaussian distributions. This allows us to perform likelihood ratio testing to effectively discriminate between intrusion and non-intrusion scenarios. Specifically, our approach differs from WiDetect [9] by considering the propagation characteristics of WiFi signals in another way, as illustrated in Fig. 1. WiFi signals can be divided into three components: low correlation components from static reflectors or LOS path, direct high correlation components introduced by in-room motions, and indirect high correlation components from out-of-room motions. Matching ACF distributions with this more practical model enables RoFi to achieve robust intrusion detection while reducing out-of-room interference.

The rest of this paper is organized as follows. We derive the estimator in Section 2 and present experimental results in

Section 3. Finally, we conclude this paper in Section 4.

2. METHOD

In this section, we model CSI measurements as statistical distributions to differentiate intrusion from non-intrusion scenarios. By analyzing the ACF of CSI power, we transform the intrinsic differences into likelihood ratio values, thereby deriving a new detection statistic. First, we introduce CSI. Next, we examine the ACF distributions before and after intrusion events. Finally, we employ the likelihood ratio method to obtain the intrusion detection statistic.

2.1. Problem Statement

For commercial WiFi devices, CSI characterizes the attenuation and phase shift introduced by WiFi signal propagation. Considering multipath effects, the measured CSI at time t is [16]:

$$H(f,t) = \sum_{l} h_l(t) e^{-j2\pi \frac{d_l(t)}{\lambda}} + n(t), \tag{1}$$

where $h_l(t)$ denotes the complex attenuation of the l^{th} path, λ is the signal wavelength, $d_l(t)$ is the path length, and $n(t)$ is random noise. The power of CSI obtained from the devices can be denoted as [9]:

$$G = |H(f,t)|^2 = \mu + \varepsilon, \tag{2}$$

where μ is the true CSI power and ε is the power introduced by noise. Then the ACF of the CSI power $\hat{\rho_G}$ can be defined as [9]:

$$\hat{\rho_G} = \frac{\sum_{t=k+1}^{N} \left(G_t - \bar{G}\right)\left(G_{t-k} - \bar{G}\right)}{\sum_{t=1}^{N} \left(G_t - \bar{G}\right)^2}, \tag{3}$$

where $\bar{G}$ is the mean of CSI power.

The ACF measures the correlation between CSI data at different time lags, helping understand historical data's influence on current values. If the CSI sequence is stationary, it would fit the Auto Regression Moving Average (ARMA) model. According to the theory outlined in [15], the ACF converges to a normal random sequence as the sample size approaches infinity.

When no intrusion occurs, WiDetect clearly demonstrates that the CSI power remains stationary, with the mean of the ACF trending towards zero. However, it only establishes the distribution of the ACF in scenarios without intrusion, neglecting to consider the distribution of the ACF during an intrusion. This theoretical oversight has limited its performance in complex scenarios.

2.2. Assumptions and Analysis

Building upon the identified limitations of WiDetect, we recognize the potential for improvement and introduce our approach. First, we consider a simple intrusion case: if an intruder keeps his activities within a fixed space like rotating around, we suppose the CSI power is also stationary, which is confirmed by an Augmented Dickey Fuller (ADF) Test [17] with 0.1 significance level on 3000 samples. Applying this to wider scenarios, if the sampling rate is sufficiently high, we can extract a short CSI segment that approximates the motion of the intruder in the same location. Thus, we derive the distribution of ACF theoretically, that is, ACF of CSI power follows Gaussian distributions $N(\mu(t), \sigma^2(t))$ with intrusion and $N(0, \sigma^2)$ without intrusion where $\mu(t) \in (0,1)$, respectively.

2.3. Detector Design

It is obvious that the mean value of ACF in the intrusion case is higher than the mean value of ACF in the non-intrusion case. This is due to the fact that the intrusion process is accompanied by motion, which inevitably causes an increase in signal correlation. Therefore, the problem of intrusion detection turns into a discrimination of which distribution fits ACF better. Thus, we take a likelihood-ratio test, which has the highest power in early hypothesis testing problems[18]. Letting f_0 and f_1 be their probability density functions of two distributions above, suppose the $\sigma^2(t)$ is fixed in a short period and the likelihood ratio is:

$$l = \frac{f_0}{f_1} = \frac{\frac{1}{\sqrt{2\pi}\sigma} e^{-x^2}}{\frac{1}{\sqrt{2\pi}\sigma(t)} e^{-(x-\mu(t))^2}} = \frac{\sigma(t)}{\sigma} e^{(x-\mu(t))^2 - x^2}. \tag{4}$$

In Equation 4, σ^2 and $\sigma^2(t)$ represent the variance of the ACF distributions under non-intrusion and intrusion scenarios, respectively. They vary across different environments (e.g. different rooms, device deployments), but share similar components within the same environment. By taking the likelihood ratio, we eliminate the shared components to improve robustness of our intrusion detector against environmental variations. Taking the logarithm of the above equation and ignoring the constant, we have:

$$\log l = (x - \mu(t))^2 - x^2. \tag{5}$$

Thus, we obtain the test statistic $\log l$.

2.4. Intrusion Statistics

Since $\mu(t)$ is unknown in practical applications, we estimate it using the sample mean ACF $\bar{\mu}$. We can derive a new test statistic $-\log \hat{l}$ as follows:

$$-\log \hat{l} = x^2 - (x - \bar{\mu})^2, \tag{6}$$

Table 1. Theoretical Comparison of Different Methods. Since it is difficult to derive an analytical solution for the estimator in WiBorder, we take the square of the original statistic as an approximation instead, which follows a scaled chi-square distribution.

Methods	WiDetect	WiBorder	RoFi
Mean(non-intrusion)	$-1/n$	$2\sigma_B^2$	$2/n^2$
Variance(non-intrusion)	$1/n$	$4\sigma_B^2$	$4/n^3 - 2/n^4$
Mean(intrusion)	$\mu(t)$	$2\sigma_B^2(t)$	$\sigma^2(t)/n + \mu^2(t)$
Variance(intrusion)	$\sigma(t)^2$	$4\sigma_B^2(t)$	$(4n - 2/n^2)\sigma^4(t) + 4\sigma^2(t)\mu^2(t)$
$u_1(n \to \infty)$	Const	Const	$0(O(1/n^2))$
$u_2(n \to \infty)$	Const	Const	Const

where $\bar{\mu} = \frac{1}{n}\sum_{i=1}^{n} \rho(i)$ represents the mean of ACF, and i is the time index. In practical applications, we use the most recently calculated ACF as the variable x, that is, $x = \rho(n)$. In this way, we obtain the new test statistic $-\log \hat{l}$.

Under the assumption and derivation above, the mean of our new statistics is $E(-\log \hat{l}) = \frac{1}{n}\sigma^2(t) + \mu^2(t)$, and the variance is $Var(-\log \hat{l}) = (\frac{4}{n} - \frac{2}{n^2})\sigma^4(t) + 4\sigma^2(t)\mu^2(t)$. n represents the sample size, $\sigma^2(t)$ represents the standard variance of noise and $\mu(t)$ represents the mean introduced by the presence of an intruder. If we use the same sample size in computing the ACF as in computing our statistic, then according to the theory of WiDetect, the mean and variance of non-intrusion can be simplified to $\frac{2}{n^2}$ and $\frac{4}{n^3} - \frac{2}{n^4}$, respectively.

2.5. Theoretical Performance Analysis

We make a theoretical comparison between the proposed statistics and two state-of-the-art intrusion detection estimators WiDetect and WiBorder based on our previous analysis, and the results are shown in Table 1.

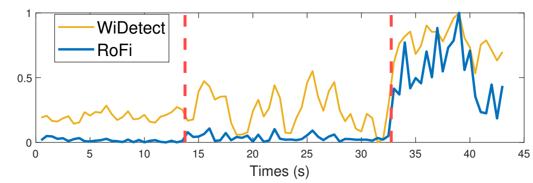
Considering different estimators having different units, we characterized the effect of the statistic by the ratio of variance to mean change $u_1 = \frac{\sigma_0^2}{\Delta\mu}$ and $u_2 = \frac{\sigma_1^2}{\Delta\mu}$, where σ_0^2 and σ_1^2 are variance of non-intrusion and intrusion scenarios. u_1 measures the degree of fluctuation of the statistic in the non-intrusion distribution, reflecting the theoretical probability of false positive and u_2 evaluates the probability of erroneous judgement of the statistic when intrusion occurs. Smaller u_1 and u_2 indicate a more stable statistic. As shown in Table 1, the limitations of all u_1 and u_2 are constant at $n \to \infty$ except for u_1 of RoFi, making RoFi theoretically superior.

3. EXPERIMENT

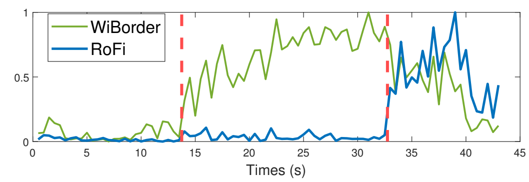
In this section, we conduct experiments to evaluate the performance of our proposed approach.

3.1. Implementation Details

The transceivers are mini PCs with Intel 5300 NIC and Linux CSI Tool [19] installed to record CSI measurements.



(a) The comparison between WiDetect and RoFi.



(b) The comparison between WiBorder and RoFi.

Fig. 2. The intrusion detection results of the three-stage experiment. During the process, the TX was placed near the door, and the RX was placed inside the room, respectively. The results are normalized to [0,1] and delineated into three distinct stages by two red dashed lines: initially an intrusion-free period, followed by movements outside the room, and finally entry into the room.

The transmitter (TX) has one antenna while the receiver (RX) has three. Each TX-RX pair has 30 subcarriers with a sampling rate of 50Hz. The TX and RX are placed in the room, with adjustable positions according to the need.

Our data encompasses diverse intrusion scenarios including normal intrusion, external patrolling, surreptitious entry, etc. Two participants with varying weights and heights separately perform identical activities to facilitate robust evaluation. Besides, to investigate the impact of perturbations such as animals, we have let a small slider (0.2 m×0.2 m×0.05 m) moved 0.5 m off the ground on a 1 m rail back and forth to simulate the animal intrusions.

3.2. Experiment Results

3.2.1. Verification Experiment

We conduct a three-stage experiment to validate our model. The experimental procedure was set as follows: Initially, no intruder was present. Subsequently, an individual moved around the door without entering. Finally, the intruder entered the room and moved randomly. Throughout the entire process, the TX remained positioned near the door, while the RX remained inside the room.

In this scenario, we compare RoFi with two baseline methods: WiDetect and WiBorder. The results depicted in Fig. 2 show that WiDetect and WiBorder are sensitive to motion, frequently misidentifying outdoor motions as intrusions. In contrast, RoFi demonstrates superior accuracy in such scenarios.

We consider non-intrusion scenarios as special motion cases, intermediate between fully unoccupied and intruded states. As modeled in Fig.1, the CSI Power comprises two main components: a low correlation signal and indirect high correlation signal. Since the former dominates energetically, the overall mean in time series of ACF is not high enough to be judged as an intrusion. However, ACF variance enables misjudgement. For such situations, the likelihood ratio test considers the overall data distribution similarity with the non-intrusion state, rather than just sample point values at one moment. This reduces noise and other artifacts, enabling more robust detection.

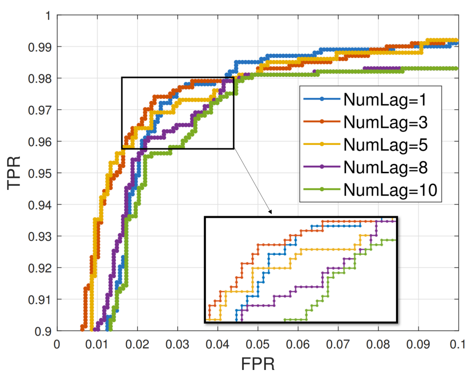


Fig. 3. ROC Curve of Different Lags.

3.2.2. *Impact of the number of ACF Lags*

The number of ACF lags significantly influences the performance of RoFi, as ACF with different lags have varying means. ACF with smaller indices tend to have larger means and variances, indicating greater sensitivity to intrusion but also greater instability. In contrast, larger lag indices exhibit more stability but reduced sensitivity. Thus, balancing ACF lag number and detection performance is crucial.

Given that the speed of intruder is estimated to be 1.3m/s on average, intruder moves about 0.078m, which approximates the wavelength of 5G WiFi signal during 3 time lags with sampling rate 50Hz. We thought 3 might be the best ACF lags, and experiment results in Fig. 3 identify with our theory.

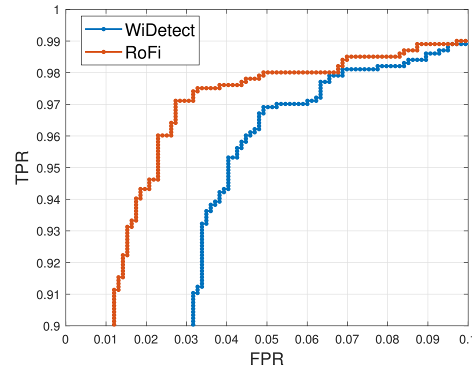


Fig. 4. ROC Curve of Different Methods.

Table 2. Comparison of Different Number of Lags.

NumLag	1	3	5	8	10
AUC	0.991	0.995	0.993	0.991	0.989
Accuracy	0.972	0.975	0.973	0.970	0.965

3.2.3. *Overall Performance*

We evaluate the total performances of different estimators, calculating different measurements as shown in Fig. 4. We can see in most cases where TPR is fixed, our RoFi holds lower FPR, indicating much less false alarms and better recognition for non-intrusion scenarios.

Table 3. Experimental Comparison of Different Methods. WiBorder* means the average of accuracy of different environments separately.

Methods	WiDetect	WiBorder	WiBorder*	RoFi
Accuracy	0.948	0.667	0.960	0.975

Table 3 lists the accuracy of different methods. The results show that our estimator outperforms others in accuracy. The performance of WiBorder in our experiment deteriorated considerably when data from different environments were aggregated, indicating the need for proper calibration before deployment.

4. CONCLUSION

In this work, we propose RoFi, a robust WiFi intrusion detector that performs well in complex situations. By modeling ACF statistically and employing likelihood ratio testing, RoFi reduces false alarms triggered by outdoor interference without requiring complex calibration. Extensive experiments demonstrate that RoFi achieves over 97.5% accuracy, indicating its advantage over existing intrusion detection methods. With excellent performance, RoFi can serve as a fundamental component for numerous smart home applications.

5. REFERENCES

[1] Fan Zhang, Albert PC Chan, and Dezhi Li, "Developing smart buildings to reduce indoor risks for safety and health of the elderly: A systematic and bibliometric analysis," *Safety Science*, vol. 168, pp. 106310, 2023.

[2] Yu-Guang Yang, Hong-Mei Fu, Shang Gao, Yi-Hua Zhou, and Wei-Min Shi, "Intrusion detection: A model based on the improved vision transformer," *Transactions on Emerging Telecommunications Technologies*, vol. 33, no. 9, pp. e4522, 2022.

[3] Simon Klakegg, Jorge Goncalves, Chu Luo, Aku Visuri, Alexey Popov, Niels van Berkel, Zhanna Sarsenbayeva, Vassilis Kostakos, Simo Hosio, Scott Savage, et al., "Assisted medication management in elderly care using miniaturised near-infrared spectroscopy," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 2, pp. 1–24, 2018.

[4] Jie Lian, Changlai Du, Jiadong Lou, Li Chen, and Xu Yuan, "Echosensor: Fine-grained ultrasonic sensing for smart home intrusion detection," *ACM Transactions on Sensor Networks*, 2023.

[5] Shuai Yang, Dongheng Zhang, Ruiyuan Song, Pengfei Yin, and Yan Chen, "Multiple wifi access points co-localization through joint aoa estimation," *IEEE Transactions on Mobile Computing*, 2023.

[6] Tianyu Zhang, Dongheng Zhang, Shuai Yang, Qibin Sun, and Yan Chen, "Wico: Robust indoor localization via spectrum confidence estimation," in *2023 IEEE Wireless Communications and Networking Conference (WCNC)*, 2023, pp. 1–6.

[7] Ganlin Zhang, Dongheng Zhang, Ying He, Jinbo Chen, Fang Zhou, and Yan Chen, "Multi-person passive wifi indoor localization with intelligent reflecting surface," *IEEE Transactions on Wireless Communications*, pp. 1–1, 2023.

[8] Shengjie Li, Zhaopeng Liu, Yue Zhang, Qin Lv, Xiaopeng Niu, Leye Wang, and Daqing Zhang, "Wiborder: Precise wi-fi based boundary sensing via through-wall discrimination," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 4, no. 3, pp. 1–30, 2020.

[9] Feng Zhang, Chenshu Wu, Beibei Wang, Hung-Quoc Lai, Yi Han, and KJ Ray Liu, "Widetect: Robust motion detection with a statistical electromagnetic model," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 3, pp. 1–24, 2019.

[10] Sameera Palipana, David Rojas, Piyush Agrawal, and Dirk Pesch, "Falldefi: Ubiquitous fall detection using commodity wi-fi devices," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 4, pp. 1–25, 2018.

[11] Hao Wang, Daqing Zhang, Yasha Wang, Junyi Ma, Yuxiang Wang, and Shengjie Li, "Rt-fall: A real-time and contactless fall detection system with commodity wifi devices," *IEEE Transactions on Mobile Computing*, vol. 16, no. 2, pp. 511–526, 2017.

[12] Dongheng Zhang, Yang Hu, Yan Chen, and Bing Zeng, "Breathtrack: Tracking indoor human breath status via commodity wifi," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 3899–3911, 2019.

[13] Youwei Zeng, Dan Wu, Jie Xiong, Enze Yi, Ruiyang Gao, and Daqing Zhang, "Farsense: Pushing the range limit of wifi-based respiration sensing with csi ratio of two antennas," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 3, no. 3, pp. 1–26, 2019.

[14] Xuecheng Xie, Dongheng Zhang, Yadong Li, Jinbo Chen, Yang Hu, Qibin Sun, and Yan Chen, "Robust respiration sensing with wifi," in *2023 IEEE Wireless Communications and Networking Conference (WCNC)*, 2023, pp. 1–6.

[15] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung, *Time series analysis: forecasting and control*, John Wiley & Sons, 2015.

[16] Dongheng Zhang, Xiong Li, and Yan Chen, "Pushing the limit of phase offset for contactless sensing using commodity wifi," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 8303–8307.

[17] Yin-Wong Cheung and Kon S Lai, "Lag order and critical values of the augmented dickey–fuller test," *Journal of Business & Economic Statistics*, vol. 13, no. 3, pp. 277–280, 1995.

[18] Jerzy Neyman and Egon Sharpe Pearson, "Ix. on the problem of the most efficient tests of statistical hypotheses," *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, vol. 231, no. 694-706, pp. 289–337, 1933.

[19] Daniel Halperin, Wenjun Hu, Anmol Sheth, and David Wetherall, "Tool release: Gathering 802.11 n traces with channel state information," *ACM SIGCOMM computer communication review*, vol. 41, no. 1, pp. 53–53, 2011.